jcls. Journal of Computational Literary Studies

OPEN ACCESS

conference version

From Review to Genre to Novel and Back

An Attempt To Relate Reader Impact to Phenomena of Novel Text

Marijn Koolen[1]
Joris J. Van Zundert[2]
Eva Viviani[3]
Carsten Schnober[3]
Willem Van Hage[3]
Katja Tereshko[2]

1. DHLab, Humanities Cluster, Amsterdam, The Netherlands.
2. Computational Literary Research, Huygens Insitute, Amsterdam, The Netherlands.
3. eScience Center, Amsterdam, The Netherlands.

Citation
Marijn Koolen, Joris van Zundert, Eva Viviani, Carsten Schnober, Willem van Hage, and Katja Tereshko (2024). "From Review to Genre to Novel and Back. An Attempt To Relate Reader Impact to Phenomena of Novel Text". In: *CCLS2024 Conference Preprints* 3 (1). 10.26083/tuprints-00027398

Date published 2024-05-28
Date accepted 2024-04-04
Date received 2024-01-25

Keywords
reader impact, literary novels, genre, topic modeling

Note
This paper has been submitted to the conference track of JCLS. It has been peer reviewed and accepted for presentation and discussion at the 3rd Annual Conference of Computational Literary Studies at Vienna, Austria, in June 2024.

Abstract. We are interested in the textual features that correlate with reported impact by readers of novels. We operationalize impact measurement through a rule-based reading impact model and apply it to 634,614 reader reviews mined from seven review platforms. We compute co-occurrence of impact-related terms and their keyness for genres represented in the corpus. The corpus consists of the full text of 18,885 books from which we derived topic models. The topics we find correlate strongly with genre, and we get strong indicators for what key impact terms are connected to which genre. These key impact terms gives us a first evidence-based insight into genre-related readers' motivations.

1. Introduction

Already Aristotle noted the reciprocal relations between an author, the text the author creates, and the response from an audience to the text. This fundamental model of rhetorical poetics has remained relevant throughout the ages (cf. e.g. Abrams 1971; Warnock 1978). The dynamics of the relations between author, text, and reader have been heavily theorized and fiercely debated (cf. e.g. Hickman 2012; Wimsatt 1954). But if there is no lack of theory, it appears to be much harder to gain empirical insights into these relations, though not for lack of trying by practitioners in such fields as empirical and computational literary studies (e.g. Fialho 2019; Loi et al. 2023; Miall and Kuiken 1994). One effect of the immense success of the World Wide Web and softwarization and digitization of societies and their cultures (Berry 2014; Manovich 2013) is the availability of large collections of online book reviews and digital full texts from novels published as ePubs. This allows us to apply NLP techniques and corpus statistics to get empirical data on the relations between text and reader that until now could only be theorized or anecdotally evidenced. At the same time, we should acknowledge that it is no panacea for the problem of empirical observations in literary studies. Not just because of the inherent biases (Gitelman 2013; Prescott 2023; Rawson and Muñoz 2016), or the almost complete lack of demographic and social signals in the data, but also because of the difficulties still involved in establishing which concrete signal in novels relates to what type of reaction for which type of reader. This is where we focus our research: we attempt to establish which concrete features of online reviews correlate to which concrete signals in the text

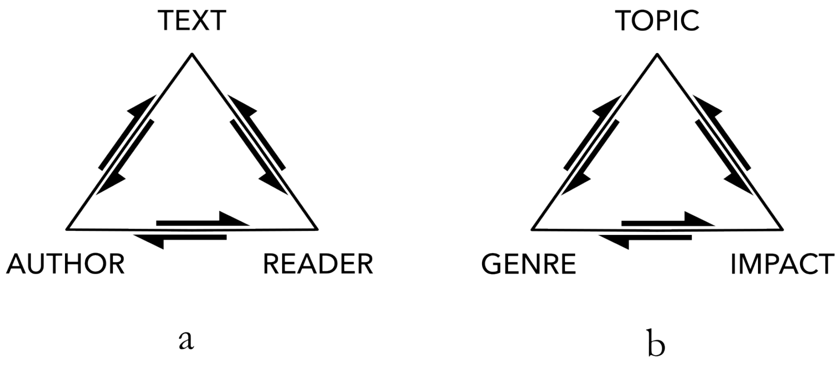


Figure 1: Classic rhetorical model (a) and our operationalization of the text–reader relation (b).

of fiction novels.

In a theoretical sense we are concentrating on the right hand side of the classical rhetorical triangle (cf. Figure 1a) and operationalize the dynamic between text and reader as another triangular relationship between *impact*, *topic*, and *genre*. With "impact" (and the commensurate "reading impact") we designate expressions of reader experiences identified by some evidence based method (e.g. as reader impact constituents researched by Koolen et al. (2023)). We apply the reader impact model to assign concrete terms to types of reading impact. The concrete text signal that we correlate this impact with are topics mined from a corpus of novels. (As an aside we note that these topics are not to be confused with themes, motives, or aboutness in a literary studies sense, as we will explain later.) A meta-textual property, genre, forms the third measurable aspect of the triangular relationship (see Figure 1b).

Concretely, we link topic models of 18,885 novels in Dutch (original Dutch and translated to Dutch) with the reading impact expressed in 130,751 Dutch online book reviews. We want to know if there is a relationship between aspects of topic in novels, their genre, and the type of impact expressed by readers in their reviews. We extracted expressions for three types of reading impact from the reviews using the previously developed Reading Impact Model for Dutch (Boot and Koolen 2020). The three types of reading impact that we discern are: "general affective impact" which expresses the overall evaluation and sentiment regarding a novel; "narrative impact", which relates to aspects of story, plot, and characters; and finally "stylistic impact" related to writing style and aesthetics.

We expect that topics in fiction are related to genre. As there is no authoritative source for genre of a novel, nor some general academic consensus about what constitutes genre, we make use of the broad genre labels that publishers have assigned to each published book. Analogous to Sobchuk and Šeļa 2023, p.2, who define genre as "a population of texts united by broad thematic similarities", we clustered these genre labels into a set of nine genres. These thematic similarities might be revealed in a topical analysis, e.g. crime novels containing more crime-related topics and romance novels containing more topics related to romance and sex. However, for some genres it might be less obvious whether they are related to topic. For instance, what are the topics one would expect in literary fiction?

It is important to note that, although the name *topic modelling* suggests that what is modelled is *topic*, most topic modelling approaches discern clusters of frequently co-occurring words, regardless of whether they have a topical connection or not (in the classical sense of "aboutness" in library science). Clusters of words may also reveal a different type of connection, e.g. words from a particular stylistic register. In that sense, genres with less clear thematic similarities may be associated with certain stylistic registers, or any other clustering of vocabulary. Different genres may also attract different types of readers and therefore different types of reviewers, who use different terminology and pay attention to different aspects of novels. It is also plausible that the language and topic of a novel influences how readers write about them in reviews. A novel written in a particularly striking poetic style may consciously or subconsciously lead readers to adopt some of its poetic aspects and register in how they write about their reading experiences. Similarly, topics in novels may be associated with what reviewers choose to mention, again, consciously or subconsciously. A novel on the atrocities of war or on the pain of losing a loved one may lead a reviewer to mention feeling sympathy or sadness during reading, while a story about friendship and betrayal might prompt reviewers to describe their anger at the actions of one of the characters.

Thus, it is clear that the relationship between the three elements – topic, genre and impact – is complex and reciprocal, as expressed in Figure 1b. Our challenge is, of course, to computationally investigate and understand this relationship utilizing the large numbers of full-text novels from different genres and corpora of hundreds of thousands of reviews. We subdivide this overarching aim into several more concrete research questions, namely:

- How are topic and impact related to each other? Do books with certain topics lead to more impact expressed in book reviews? Do different topics lead to different types of impact?
- How are genre and impact related to each other? Do books of different genres lead to different types of impact? Do reviews of different genres use different vocabulary for expressing the same types of impact?
- How are topic and genre related to each other? Are certain topics more likely in some genres than in others?

This paper makes three main contributions to our ongoing research. The first is that it contributes to our understanding of the reading impact model, and through it, of the language of reading impact. We formalize the ability to tell genres apart using the *keyness* of impact terms. Thus, we now have quantitative support to argue that certain impact terms are strongly connected to certain genres and less to others. Second, we find that the topics from novels can be clustered into broader themes that lead to distinct thematic profiles per genre. There is a clear relation between impact terms and genre, but not between impact terms and topic or theme. In the discussion at the end we elaborate on this and provide possible explanations for this finding. The third contribution is the insight that the key impact terms per genre give an indication of the motivation of readers to read a book and how the reading experience relates to their expectations.

2. Background

We are interested in what kind of impression novels leave with their readers. Can we measure this so-called "impact" and how does it relate to features of the actual novel texts? Several studies have tried to link success or popularity of texts to features of those texts. Some studies have related pace, in the sense of how much distance the same length of texts covers in a semantic space, to success; finding that success correlates with higher pacing of narrative (Toubia et al. 2021, Laurino Dos Santos and Berger 2022). It has been argued that songs of which lyrics deviates form a genre's usual pattern tend to be more popular (Berger and Packard 2018). Other work relates topic models to surveyed ratings of literariness suggests the same for fiction novels (Cranenburgh et al. 2019). Moreira et al. apply "sentiment arc features [...] and semantic profiling" with some success to predict ratings on Goodreads (Moreira et al. 2023). Taking the number of Gutenberg downloads as a proxy for success Ashok et al. (2013) reach 84% accuracy in predicting popularity based on learning low level stylistic features of the text of novels. Van Zundert et al. (2018) use sales numbers as a proxy for popularity in an machine learning attempt to predict success, concluding that the theme of masculinity is at least one major driver of successful fiction.

Common to all these studies is that they target some proxy of success or popularity: Goodreads ratings, sales numbers, download statistics, and so forth. However, to our knowledge no research has tried to link concrete features of fiction narratives to textual features of reviews from readers. We seek to uncover if there is such a relation and if it may be meaningful from a literary research perspective. In our present study we apply a heuristic model for impact features (Boot and Koolen 2020) to a corpus of 600,000+ reader reviews mined from several online review platforms. We attempt to relate collocations of impact related terms to genre. Advancing previous research on genre and topic models (Van Zundert et al. 2022) our contribution in this paper is to examine how collocated impact terms relate to genre and genre to topic models of novels, thus offering a first insight into the relation between topics (understood in terms of topic model) and reader reported impact measures. Such work needs to take into account the plethora of problems that surround the application of topic models to downstream tasks. This concerns topics content wise, which is to say that topic models in contrast to their name do not often express much topical information. Rather they may be connected to meta-textual features, such as author (Thompson and Mimno 2018), genre (Schöch 2017), or structural elements in texts (Uglanova and Gius 2020).

Our current contribution leans more to the side of data exploration than to the side of offering assertive generalizations. We are interested in empirically quantifying the impact that the text of novels has on readers. Any operationalization of this research aim necessarily involves many narrowing choices and, at least initially, the audacious naivety to ignore the stupefying complexity of social mechanisms to which readers are susceptible and thus the mass of confounding text external factors that also drive reader impact. In our setup we assume that there are at least some textual features, such as style, narrative pace, plot, character likability, that may be measured and that can be related to reader impact. We further assume that book reviews scraped from online platforms do serve as a somewhat reliable gauge to measure reader impact. We make these cautioning statements not just proforma, but because we know that our information is selective, biased, and skewed. Thanks to the stalwart experts of the Dutch National Library we do have for our analysis the full text of 18,885 novels in Dutch (both translated and of Dutch origin). We also have 634,614 online reviews, gathered by scraping for platforms such as Goodreads, Hebban[1], and so forth. This corpus is biased. Romance novels comprise only about 3% of the corpus of full texts. This is in stark contrast to its undisputed popularity (cf. Regis 2003, p. xi: "In the last year of the twentieth century, 55.9% of mass-market and trade paperbacks sold in North America were romance novels"). If our book corpus is skewed, our review data is even more so: only 1% of reviews pertain to novels in the romance genre. Obviously we attempt to balance our data with respect to genre and other properties for analysis. Yet, we should remind ourselves of the limited representativeness of our data, which necessitates modesty as to generalizing results. Hence, what follows is more offered as data exploration than as pontification of strong relations.

1. See https://www.hebban.nl/.

3. Data and Method

Our corpus of 18,885 books consists of mostly fiction novels and some non-fiction books in the Dutch language (both originally Dutch and translated). The review corpus boasts 634,614 Dutch book reviews. Obviously we do not have reviews for each book, nor does the set of books fully cover the collection of reviews, but we have upward of 10k books with at least one review.

3.1 Preprocessing

Both books and reviews are parsed with Trankit (Nguyen et al. 2021). Reading impact is extracted from the reviews using the Dutch Reading Impact Model (DRIM) (Boot and Koolen 2020).

Topic Modelling For topic modelling of the novels we use Top2Vec (Angelov 2020), and created a model with whole books as documents. We apply multiple filters to select terms that signal topic. Following the advice from previous work (Sobchuk and Šeļa 2023; Uglanova and Gius 2020; Van Zundert et al. 2022), we focus on content words and select only nouns, verbs, adjectives and adverbs and remove any person names identified by the Trankit NER tagger. Our assumption is that person names have little to no relationship with topic, but are strong differentiating terms that tend to cluster parts of books and book series with recurring characters. Names of locations can have a similar effect, but, at least where the setting reflects the real world, we argue that this setting aspect of stories is more meaningfully related to topic. The book corpus contains 1,922,833,614 tokens including all punctuation and stop words. After filtering, 826,226,855 tokens remain. The next filter is a frequency filter. We remove terms that occur in fewer than 1% of documents or in more than 50% of documents. This leaves 190,607,470 tokens, which is 23% of all content words and just under 10% of the total number of tokens[2]. Books have a mean (median) number of 42,959 (37,940) *content* tokens. The number of tokens is a Poisson distribution, therefore left-skewed, with 68% (corresponding to data within 1 standard deviation from the mean) of all books having between 17,509 and 63,418 tokens. This shows that the books have a high variation in length, but the majority books have a length within a single order of magnitude. After filtering on document frequency, the mean (median) number of tokens is 9,979 (8,325), with 68% having between 3,847 and 14,992 tokens.

Reading Impact Modelling The DRIM is a rule-based model and works at the level of sentences. It has 275 rules relating to impact in four categories: *Affect*, *Aesthetic* and *Narrative* impact, and *Reflection*. Both *Aesthetic* and *Narrative* impact are sub-categories of *Affect*, so rules that identify expressions of the sub-categories are also considered expressions of *Affect* (Boot and Koolen 2020). The rules for *Reflection* were not validated (see Boot and Koolen 2020) so we exclude *Reflection* from our analysis. For our analysis of topic, we expect that *Narrative* is the most directly related category, but we also include general *Affect* in our analysis. Expressions identified by the model consist of at least an impact word or phrase, such as “spannend” (*suspenseful*).[3] However, many rules require there to be a book aspect term as well. For instance, the evaluative word “goed” (*good*) by itself can refer to anything. To be considered part of an impact expression it needs to co-occur in one sentence with a word in one of the book aspect categories, e.g. a style-related word

2. Experiments with using different frequency ranges for filtering suggests that the topic modelling process is relatively insensitive with regards to the upper limit. I.e. using 50%, 30% or 10% results in roughly equal numbers of topics that show the same relationship with book genre (see Section 4.1.1 and the following notebook: https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact/blob/main/notebooks/topic_and_genre.ipynb
3. For all Dutch terms we will consistently provide English translation in italics between parentheses.

like "geschreven" (*written*) to be an expression of *Aesthetic* impact, or a narrative-related word like "verhaal" (*story*) or "plot" to be an expression of *Narrative* impact.

The DRIM identified 2,089,576 expressions of impact in the full review dataset. To identify the key impact terms per genre, we use the full review dataset with all 2.1M impact expressions. To make a clearer distinction between impact expressions of generic affect and affect specific to narrative or aesthetics, we consider as *Affect* only those expressions that are not also categorized as *Narrative* or *Aesthetic*. Of the 2,089,576 expressions, there are 667,672 expressions for *Aesthetic* impact, 690,184 for *Narrative* impact and 731,720 for generic *Affect*.

3.2 Connecting Books and Reviews

A crucial step in relating topic in fiction to reading impact expressed in reviews, we need to connect the books to their corresponding reviews. For this, we rely mostly on ISBN[4] and author and book title. Note that a particular work may be connected to multiple ISBNs, for instance when reprints or new editions are produced for the same work with a different ISBN. Many mappings between reviews and books, and between multiple ISBNs of the same work were already made by Boot 2017 and Koolen et al. 2020, for the Online Dutch Book Response (ODBR) dataset of 472,810 reviews. We added around 160,000 reviews from Hebban to the ODBR set. To find ISBNs that refer to the same work, we first queried all ISBNs found in reviews using the SRU[5] service of the National Library of the Netherlands. This SRU service gives access to the combined catalog of Dutch libraries and in many cases links multiple editions of the same work with different ISBNs. Using author and title we resolved another number of duplicated works with different ISBNs. We then mapped all ISBNs of the same work to a unique work ID and linked the reviews via the ISBNs they mention to these work IDs. There are 125,542 distinct works reviewed by the reviews in our dataset. Of the 18,885 books for which we have ePubs, there are 10,056 books with at least one review in our data set. Altogether these 10,056 unique works are linked to 130,751 reviews.

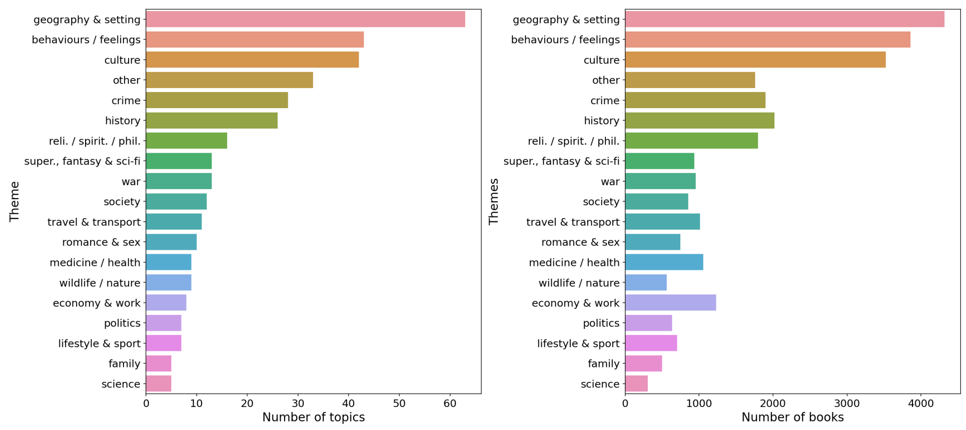
3.3 Connecting Impact and Topic Data

Our goal was to have a comprehensive mapping of the most relevant topics of works to their reviews, the latter analyzed via the DRIM. To create this dataset, we needed to connect the expressions of impact to the topics in our book dataset. To do so, we took the top five dominant topics of each book[6], and linked those topics to the impact expressions in the reviews of the books for that topic. This resulted in a dataset whereby each entry links specific reviews to the top 5 dominant topics for every book.

The Top2Vec model gave us a total of 228 topics. We attempted to label each topic with a distinct content label, but found that many topics are thematically very similar, capturing many of the same elements. Therefore, we manually assigned each topic to one or more of 19 broader themes: 1. *geography and setting*, 2. *behaviors/feelings*, 3. *culture*, 4. *crime*, 5. *history*, 6. *religion, spirituality and philosophy*, 7. *supernatural, fantasy and sci-fi*, 8. *war*, 9. *society*, 10. *travel and transport*, 11. *romance and sex*, 12. *medicine/health*, 13. *wildlife/nature*, 14. *economy and work*, 15. *lifestyle and sport*, 16. *politics*, 17. *family*, 18. *science*, 19. *other*. We provide the number of topics grouped per

4. International Standard Book Number, see: https://en.wikipedia.org/wiki/ISBN.
5. Search and Retrieval by URL, see: https://en.wikipedia.org/wiki/Search/Retrieve_via_URL.
6. Topc2Vec creates topics by clustering the document vectors and taking the centroid of each cluster as the topic vector. We computed the cosine similarity between the document vector (representing the book) and the topic vectors, and selected the top five closest (i.e., most similar) topics to each book.

Figure 2: The number of topics and books per theme.



theme in Figure 2[7].

We provide the full list of topics, themes and their respective words in our code repository[8].

3.4 Book Genre Information

For genre information about books, we use the Dutch NUR classification codes assigned by publishers. As NUR was designed as a marketing instrument to determine where books are shelved in bookshops, publishers can choose codes based not only on the perceived genre of a book but also on marketing strategies related to where they want a book to be shelved to find the biggest audience. Some NUR codes refer to the same or very similar genres. E.g. codes 300, 301, and 302 refer respectively to *general literary fiction*, *Dutch literary fiction*, and *translated literary fiction*, which we group together under *Literary fiction*. Similarly, we group codes 313, 330, 331, 332, and 339 under *Suspense* novels, as they all refer to types of suspense, i.e. *pockets suspense*, *general suspense novels*, *detective novels* and *thrillers* respectively. In total, we select 19 different NUR codes and map them to 9 genres. All remaining NUR codes in the fiction range (300-350) we map to *Other fiction* and the rest to *Non-fiction*. The full mapping is available in our code repository[9].

3.5 Keyness Analysis on Impact Terms

The goal of this analysis is to determine (i) *which* words readers use in their reviews to describe the impact of a particular book, and (ii) how *characteristic* these words are for a particular genre, compared to another genre. A good candidate to measure both (i) and (ii) is keyword analysis, or keyness (Dunning 1994; Gabrielatos 2018; Paquot and Bestgen 2009).

There is ample literature comparing different keyness measures (Culpeper and Demmen 2015; Du et al. 2022; Dunning 1994; Gabrielatos 2018; Lijffijt et al. 2016), finding that no single measure is perfect.

A commonly used measure is G^2, which identifies *key* terms that occur statistically significantly more or less often in a target corpus (the reviews for a particular genre) compared to a reference

7. Note that in this paper "theme" should not be taken to coincide with the literary studies sense of theme. Rather we use the term "theme" to clearly distinguish between the topics as identified by Top2Vec and their clustering as done by us.
8. See `https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact/blob/main/data/topic_labels.tsv`.
9. See `https://anonymous.4open.science/r/jcls-2024-topic-genre-impact-EB46/data/nur_genre_map.md`.

corpus (reviews of one or more other genres).

Lijffijt et al. (2016) showed that Log-Likelihood Ratio (G^2, Dunning 1994) and several other frequency-based bag-of-words keyness measures suffer from excessively high confidence in the estimates because these measures assume samples to be statistically independent, but words in a text are not independent of each other. Du et al. (2022) compare frequency-based and dispersion-based measures for a downstream task (text classification) to show that for identifying key terms in a sub-corpus compared to the rest of the corpus, dispersion-based measures are more effective.

To compare the dispersion of a word or phrase in a target corpus to its dispersion in a reference corpus, Du et al. (2021) introduce *Eta*, which is a variant of the *Zeta* measure by Burrows (2006).

They find that Eta Du et al. 2021 and Zeta Burrows 2006 are among the most effective measures. Both Eta and Zeta compare document proportions of keywords. The former uses Deviation of Proportions (DP) Gries 2008 which computes two sets of proportions. The first are the proportions that the lengths of documents represent with respect to the total number of words in a corpus (e.g. the set of reviews for books of a specific genre) as an expected distribution of proportions of keywords. The second is the set of observed proportions of a keyword across a corpus with respect to the total corpus frequency of that keyword. There are two problems with using DP for keyness of impact terms. The first is that some impact terms do not occur in any of the reviews of a specific genre. In such cases, the observed proportions are not properly defined (a proportion of zero is not well-defined), so DP cannot be computed. The second is that the frequency distribution of impact terms in reviews is extremely skewed (84% of all impact terms in reviews have a frequency of 1, 13% occur twice and the remaining 3% occur three or four times). Although longer reviews have a higher a priori probability of containing a specific impact term than shorter reviews, the frequency distribution of individual impact terms behaves more like a binomial distribution, so length-based proportions are not an appropriate measure of keyness.

Because of this, we instead measure dispersion using *document frequencies* (the number of reviews for a book genre in which an impact term occurs) to compute the *document proportion* (the fraction of reviews for a book genre in which an impact term occurs at least once). This gives document proportion $docP(t, G)$ per impact term t and genre G, with the absolute difference *Zeta* between two genres defined as $Zeta(t, G_1, G_2) = abs(docP(t, G_1) - docP(t, G_2))$.

To illustrate this approach, we compare the document proportions per genre of the impact terms "stijl" (style) and "schrijfstijl" (writing style). The former has the highest document proportion for reviews of *Literary fiction* (occurring in 3.7% of reviews) and least in those of *Non-fiction* (1.2%), resulting in $Zeta = 0.037 - 0.012 = 0.025$. The latter is most common in reviews of *Romanticism* (14.6%) and least common in those of *Non-fiction* (2.0%), giving $Zeta = 0.146 - 0.02 = 0.126$.

4. Results

4.1 Topic and Genre

Van Zundert et al. (2022) found that the topics identified with Top2Vec are strongly associated with genre as identified by publishers. Similarly, Sobchuk and Šeļa 2023 find that Doc2Vec – which is used by Top2Vec to embed the documents in the latent semantic space in which topic vectors are identified – is more effective at clustering books by genre than the topic modeling technique LDA (Blei et al. 2003).

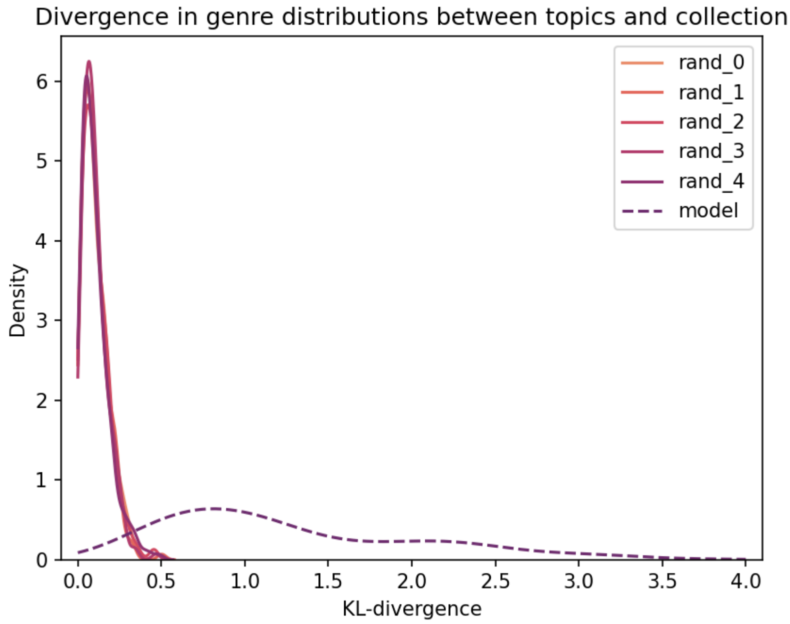


Figure 3: The KL-divergence between the genre distribution per topic and that of the collection for the topic model as well as for five random shufflings of genre labels using the same books per topic.

4.1.1 Genre Distribution per Topic

To extent the findings of Van Zundert et al. 2022, we first quantitatively demonstrate that there is a relationship between topic and genre. Each topic is associated with a number of books and thereby with the same number of genre labels. From eyeballing the distribution of genre labels per topic, it seems that for most topics, the vast majority of books in that topic belong to a single genre. But the genre distribution of the entire collection is also highly skewed, with a few very large genres and many much smaller genres. So perhaps the skew in most topics resembles the skew of the genre distribution of the collection.

To measure how much the genre distribution per topic deviates from that of the collection, we compute the KL-divergence between the two distributions. This gives a set of 228 deviations from the collection distribution.

But whether these deviations are small or large is difficult to read from the numbers themselves. For that, we should compare them against a random shuffling of the book genres across books (while keeping the books assigned per topic stable). For large topics (with many books), a random shuffling should have a genre distribution close to that of the collection. For small clusters, the divergence will tend to be higher.

We create five alternative clusterings with books randomly assigned to topics with the same topic size distribution as established by the topic model. The distribution of the 228 KL-divergence scores per model (five random and one topic model) are shown in Figure 3. The five random models have almost identical distributions concentrated around 0.1 with a standard deviation of around 0.075 and a max of around 0.5. The genre distribution of the topic model is very different, with a median score of 1.06 and more than 75% of all scores above 0.68.

From this quantitative analysis, it is clear that there is a strong relationship between topic and genre.

4.1.2 Thematic Distribution per Genre

Next, we perform a qualitative analysis of the topics and their relationship to genre.

The distribution of topic themes per genre is shown in Figure 4 in the form of radar plots. The genres show distinct thematic profiles. Literary fiction scores high on the themes of *Culture*, *Geography & setting* and *Behaviors & feelings*, which is perhaps not surprising. Non-fiction scores high on *Religion, spirituality, and philosophy*, *Medicine & health*, *Economy & work*, and *Behaviors & feelings*, which are themes that few fiction genres score high on.

In Children's fiction, there is relatively little use of the geographical aspect of setting, especially compared to other fiction genres. That is, it seems that children's novels make little explicit reference to geographical places. They score high on *behaviors and feelings* and moderately high on *Culture*, *Family* and *Supernatural, fantasy & sci-fi*. The main difference between Children's fiction and Young Adult is that the letter scores higher on *Supernatural, fantasy and sci-fi*. On the former theme, Young Adult strongly overlaps with Fantasy novels. Young Adult also adds in a bit of *Romance and sex*. These observations suggest that Children's fiction and Young Adult by and large treat the same themes but against different 'backgrounds'. Children's fiction is about behaviors and feelings against a backdrop made up of culture and family. Young adult does practically the same, but adds supernatural, fantasy, and sci-fi elements to the story, and opens the stage for some romantic behavior.

If one would want to hazard a guess at reader development, it would almost seem as if young readers are invited to pre-sort on the major themes of grown-up literature where *Romance* amplifies the romance and sex encountered in *Young adult* books, while *Literary fiction* and *Literary thrillers* amplify motifs of culture, setting, and crime, and *Fantasy* caters to the interest in the supernatural developed through Young adult fiction. Much more research would be needed, however, to substantiate such a pre-sorting effect. In any case, Romanticism scores high on *Romance and sex* and has medium scores for *Culture* and *Geography and setting*, while Suspense novels score high on *Crime*, and have medium scores for *Geography and setting* and *War*.

We expect that many of these observations coincide with intuitions of literary researchers. This suggests that the grouping of topics by theme makes sense from a literary analytical perspective in any case. The findings also shows where genres overlap and where they differ. For instance, the profile for Literary fiction and Literary thriller are similar, with the main difference being the much higher prevalence of the *Crime* theme in Literary thrillers. Suspense is similar to Literary thrillers in the prevalence of *Crime* as theme, but lower scores for *Culture* and *Geography and setting*.

One of the main findings is that, for the chosen document frequency range of mid-frequency terms, there is a clear connection between topic and genre, with thematic clustering of topics leading to distinct genre profiles, but also to thematic connections between certain genres. None of this will radically transform our understanding of genre and topic, but it prompts the question how different parts of the document frequency distribution relate to different aspects of novels. From authorship attribution research we know that authorial signal is mainly found in the high-frequency range, and our work corroborates earlier findings that topics contain genre-signals in mid-range frequencies (Thompson and Mimno 2018; Van Zundert et al. 2022).

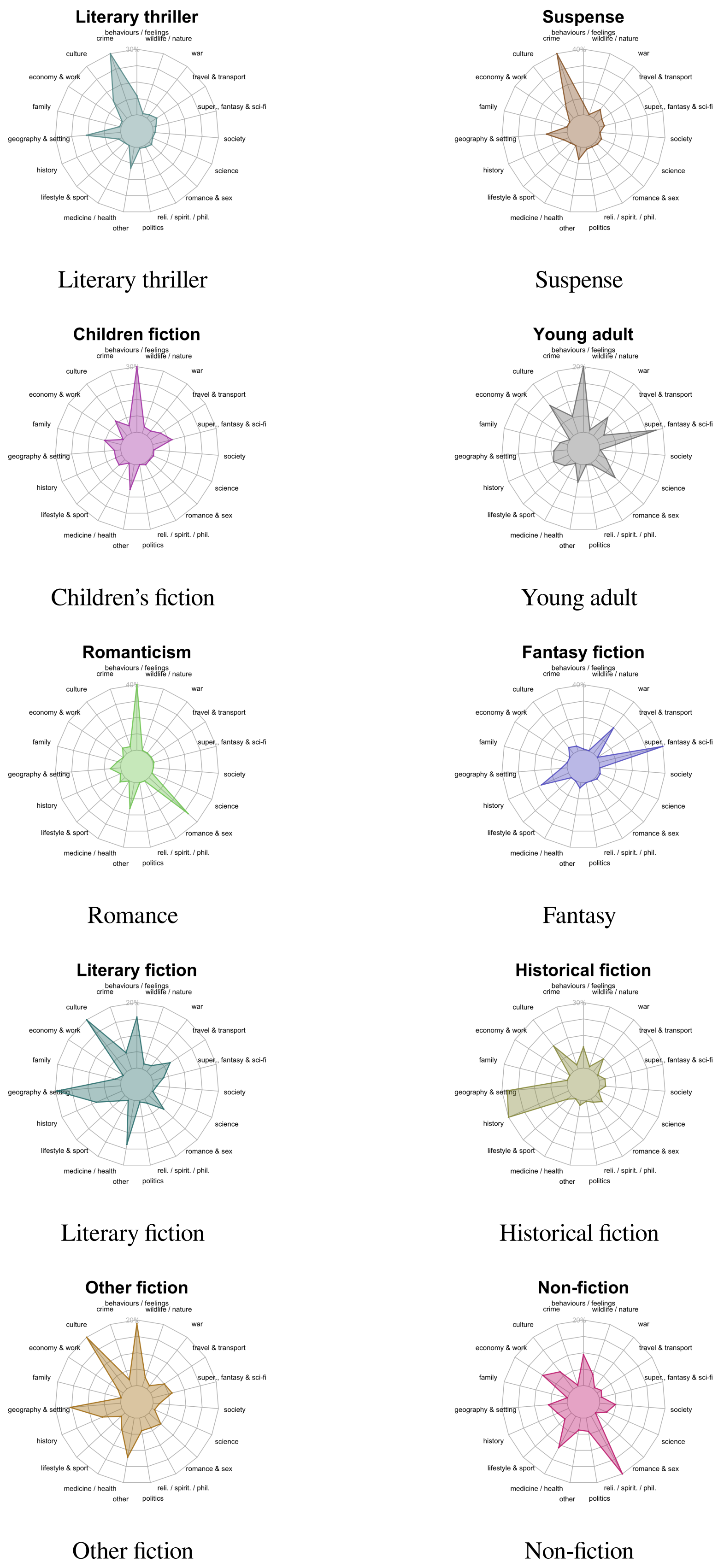


Figure 4: Radar plots showing the relative prevalence of themes in six genres, from left to right, top to bottom: *Literary thrillers*, *Suspense*, *Children's fiction* and *Young adult*, *Romance*, *Fantasy*, *Literary fiction*, *Historical fiction*, *Other fiction* and *Non-fiction*.

Table 1: Reviews per genre and mean number of reviews per book, per genre.

	Reviewed books	Reviews	Mean Reviews/book
Literary fiction	19288	200907	10.4
Literary thriller	3394	77288	22.8
Young adult	2919	30552	10.5
Children fiction	5348	27989	5.2
Suspense	6266	67990	10.9
Fantasy fiction	1571	13739	8.7
Romanticism	1291	6434	5.0
Historical fiction	556	3463	6.2
Regional fiction	472	1528	3.2
Other fiction	7260	37515	5.2
Non-fiction	26884	109158	4.1

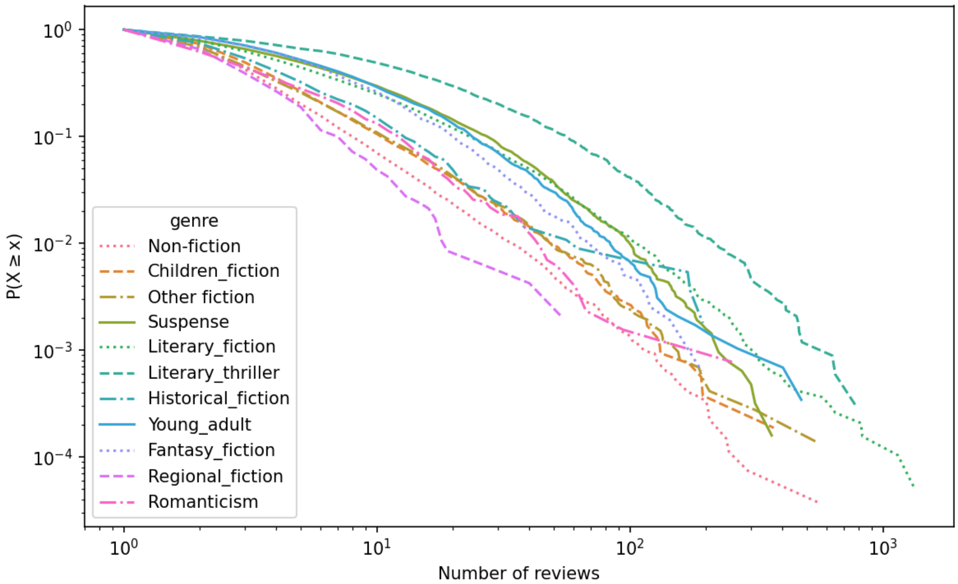


Figure 5: The cumulative distribution function of the number of reviews per book, on a log-log scale. The Y-axis shows that probability $P(X \geq x)$ that a book has at least x reviews.

4.2 Impact and Genre

4.2.1 Reviews per Genre

With the genre labels, we can count how many books in each genre have reviews in our dataset, and how many reviews they have (Table 1). It is clear that *Literary fiction* is reviewed most often, with 200,907 reviews in our dataset, followed by *Literary thrillers* and *Suspense novels*. *Literary thrillers* have the highest mean number of reviews per book. However, the distribution of the number of reviews per book is highly skewed, with a single review per book being the most likely, and having more reviews being increasingly unlikely (Koolen et al. 2020). The distributions per genre show some differences, but all are close to a power-law. The cumulative distribution function of the number of reviews per book for the different genres are shown in Figure 5, with on the Y-axis the probability $P(X \geq x)$ that a book has at least x reviews.[10]

The curves for some of the genres overlap, which makes them difficult to discern, but there are a few main insights. First, *regional fiction* and *non-fiction* have the fastest falling curves, indicating that books in these genres are the least likely to acquire many reviews. Next is a cluster of *children's fiction*, *romanticism*, *historical fiction* and *other fiction*, which tend to get a slightly higher number of reviews. Then there is a cluster of *suspense*, *literary fiction*, *young adult* and *fantasy fiction*, which tend to get more reviews than the previous cluster. And finally, clearly above the rest, is the curve of *literary thrillers*, which tend get more reviews than books in any other genre.

Thrillers are more often reviewed on the platforms that are in the review dataset. *Romance* novels have fewer reviews but are a very popular genre (Regis 2003, p. 108, see also: Darbyshire 2023). This prompts the question of whether readers of *regional* and *romance* novels have less desire to review these novels or review them on different platforms and in different ways. As there seem to be many video reviews of *romance* novels on TikTok using the tag #BookTok, this would be a valuable resource to add to our investigations. A difference in the number of reviews might be a signal of a difference in impact, but it is also plausible that different genres attract different types of readers who express their impact in different ways linguistically, using different media (e.g. text or video) on different platforms (e.g. GoodReads or TikTok). To that extent, the review dataset may be a biased representation of the impact of books in different genres. Bracketing for a moment the potential skewedness of the number of reviews per genre, and taking number of reviews as a proxy of popularity, it is also interesting to observe that popularity is apparently a commodity that is reaped in orders of magnitude.

4.2.2 Key Impact Terms per Genre

Correlations between genres First, we compare genres in terms of their impact terms through the percent difference per impact term. For each pair of genres, we compute the Pearson correlation ρ between the $\%Diff$ scores of all impact terms. A high positive correlation means that impact terms with high (low) $\%Diff$ scores in one genres, tend to also have high (low) $\%Diff$ scores in the other genre.

The correlations per impact type are shown Figure 6. For *Affect* impact terms (the top correlation table), many of the genre pairs have no correlation ($-0.25 < \rho < 0.25$). There are some weak positive and negative correlations ($0.25 < \rho < 0.50$ and $-0.50\rho < -0.25$ respectively) and

10. We show the cumulative distribution instead of the plain distribution, because it produces smoother curves and better shows the trends.

Affect

	Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	YA
Child. fic	1.00	0.44	-0.15	-0.34	-0.15	0.17	-0.11	-0.09	0.14	-0.12	0.50
Fantasy	0.44	1.00	0.09	-0.42	0.22	-0.20	-0.14	-0.13	0.05	0.24	0.60
Hist. fic	-0.15	0.09	1.00	-0.03	0.07	-0.10	-0.15	0.05	0.22	-0.11	0.18
Lit. fic	-0.34	-0.42	-0.03	1.00	-0.45	-0.18	-0.14	-0.08	-0.09	-0.36	-0.40
Lit. thrill	-0.15	0.22	0.07	-0.45	1.00	-0.38	-0.19	0.21	-0.09	0.61	0.13
Non-fic	0.17	-0.20	-0.10	-0.18	-0.38	1.00	-0.05	-0.09	-0.07	-0.37	-0.15
Oth. fic	-0.11	-0.14	-0.15	-0.14	-0.19	-0.05	1.00	-0.05	-0.08	0.01	-0.17
Reg. fic	-0.09	-0.13	0.05	-0.08	0.21	-0.09	-0.05	1.00	0.39	-0.08	0.08
Romance	0.14	0.05	0.22	-0.09	-0.09	-0.07	-0.08	0.39	1.00	-0.21	0.34
Suspense	-0.12	0.24	-0.11	-0.36	0.61	-0.37	0.01	-0.08	-0.21	1.00	-0.03
YA	0.50	0.60	0.18	-0.40	0.13	-0.15	-0.17	0.08	0.34	-0.03	1.00

Narrative

	Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	YA
Child. fic	1.00	0.60	-0.12	-0.41	-0.06	0.09	-0.13	0.02	0.07	0.07	0.46
Fantasy	0.60	1.00	-0.02	-0.41	0.09	-0.29	-0.15	-0.10	0.13	0.24	0.48
Hist. fic	-0.12	-0.02	1.00	0.05	-0.01	-0.15	-0.18	0.50	0.46	-0.33	0.24
Lit. fic	-0.41	-0.41	0.05	1.00	-0.44	-0.04	-0.18	0.01	-0.07	-0.35	-0.32
Lit. thrill	-0.06	0.09	-0.01	-0.44	1.00	-0.38	-0.25	-0.12	-0.14	0.32	0.12
Non-fic	0.09	-0.29	-0.15	-0.04	-0.38	1.00	-0.08	0.02	-0.06	-0.30	-0.22
Oth. fic	-0.13	-0.15	-0.18	-0.18	-0.25	-0.08	1.00	-0.14	-0.10	-0.00	-0.23
Reg. fic	0.02	-0.10	0.50	0.01	-0.12	0.02	-0.14	1.00	0.48	-0.30	0.20
Romance	0.07	0.13	0.46	-0.07	-0.14	-0.06	-0.10	0.48	1.00	-0.25	0.44
Suspense	0.07	0.24	-0.33	-0.35	0.32	-0.30	-0.00	-0.30	-0.25	1.00	-0.26
YA	0.46	0.48	0.24	-0.32	0.12	-0.22	-0.23	0.20	0.44	-0.26	1.00

Style

	Child. fic	Fantasy	Hist. fic	Lit. fic	Lit. thrill	Non-fic	Oth. fic	Reg. fic	Romance	Suspense	YA
Child. fic	1.00	0.16	-0.05	-0.25	-0.30	0.08	-0.07	0.13	0.43	-0.36	0.60
Fantasy	0.16	1.00	0.03	-0.43	0.49	-0.24	-0.46	0.06	0.02	0.19	0.43
Hist. fic	-0.05	0.03	1.00	-0.21	0.24	-0.06	-0.34	0.54	-0.08	0.10	0.15
Lit. fic	-0.25	-0.43	-0.21	1.00	-0.44	-0.36	0.34	-0.34	-0.18	-0.44	-0.49
Lit. thrill	-0.30	0.49	0.24	-0.44	1.00	-0.31	-0.65	0.16	-0.23	0.57	0.13
Non-fic	0.08	-0.24	-0.06	-0.36	-0.31	1.00	0.17	-0.03	-0.04	-0.25	-0.12
Oth. fic	-0.07	-0.46	-0.34	0.34	-0.65	0.17	1.00	-0.10	0.24	-0.35	-0.31
Reg. fic	0.13	0.06	0.54	-0.34	0.16	-0.03	-0.10	1.00	0.39	0.05	0.33
Romance	0.43	0.02	-0.08	-0.18	-0.23	-0.04	0.24	0.39	1.00	-0.23	0.47
Suspense	-0.36	0.19	0.10	-0.44	0.57	-0.25	-0.35	0.05	-0.23	1.00	-0.01
YA	0.60	0.43	0.15	-0.49	0.13	-0.12	-0.31	0.33	0.47	-0.01	1.00

Figure 6: Pearson correlation in the $\%Diff$ scores of impact terms between pairs of genres, for *Affect* (top), *Narrative* (middle) and *Style* (bottom).

moderate correlations ($0.50 < \rho < 0.75$ and $-0.75\rho < 0.50$). There are a few clusters of genres with high correlations in $\%Diff$ scores, signaling that some genres differ in how impact is expressed and that the DRIM is sensitive to difference between genres. The cluster of Children's fiction, Young adult and Fantasy have weak (0.44) and moderate (0.50 and 0.60) correlations with each other, suggesting that impact terms that are typical for one, are to some extent also typical for the other two. Other clusters are Literary thriller and Suspense novels, with a moderate correlation of 0.61, and Romance and Regional fiction with a moderate correlation of 0.39.

Literary fiction is the one genre with mostly weakly negative correlations, with Children's fiction (-0.34), Fantasy (-0.42), Literary thriller (-0.45), Suspense (-0.36) and Young adult (-0.40). With the remaining three genres, literary fiction has no correlation. In other words, in terms of *Affective* impact, reviews of Literary fiction uses a different register than reviews of other genres.

For *Narrative* impact, we find the same cluster of Children's fiction, Young adult and Fantasy. The cluster of Regional fiction and Romance here also contains Historical fiction, and the two clusters are linked by the moderate correlation of 0.44 between Romance and Young adult. The other genres in the two clusters have no or a negative correlation with each other. Here also the genres of Literary thriller and Suspense novels show a weak correlation (0.32), and Literary fiction has no or at most moderately negative correlations with the other genres. The top impact terms for Thrillers and Suspense novels largely overlap and contain several narrative impact terms relating to plot, e.g. "spannend" (*thrilling* or *suspenseful*), "spanning" (*suspense*), "verrassing", "verrassend" and "onverwacht" (*surprise*, *surprising* and *unexpected* respectively). For Romance and Regional fiction, the top 10 narrative impact terms almost completely overlap, with shared narrative impact terms "romantisch" (*romantic*), "ellende" (), "verdriet" (*sadness*), "levensecht" (*lifelike*), "fijn" (*nice*), "heerlijk" (*lovely*) and "nieuwsgierig" (*curious*).

Overall, there are more weak negative correlations between pairs of genres that for *Affective* impact were non-existent.

The correlations for *Style* are more different. Children's fiction no longer has a weak positive correlation with Fantasy, but it does with Romance. Children's fiction and Young adult still have a moderately positive correlation and Young adult also have weak correlations with Fantasy and Romance. The biggest shifts are for Romance, which no longer has any correlation with Historical fiction, but now has a weakly positive correlation with Children's fiction. For Literary thrillers there are several weakly and moderately negative correlations with Children's fiction (-0.30), Literary fiction (-0.44), Non-fiction (-0.31) and Other fiction (-0.65). Literary fiction is also in terms of *Style* different from almost all genres apart from Other fiction. A speculative interpretation is that Literary fiction is stylistically distinctive in a similar way to the poetry that is part of the Other fiction genre.

Compared across the different impact types then, it appears that Literary fiction as a genre induces reviews where impact is described in a vocabulary distinct from impact reported in reviews pertaining to other genres. It is tempting to conjecture that Literary fiction attracts an audience of review writers that 'know how to talk' about literature. It is very well possible that these reviewers are acutely aware of the genre of literary review and that they apply conventions of this genre in their own review writing. For now this must remain indeed conjecture as a more focused examination of the vocabulary, style, and structure of these reviews has yet to be undertaken.

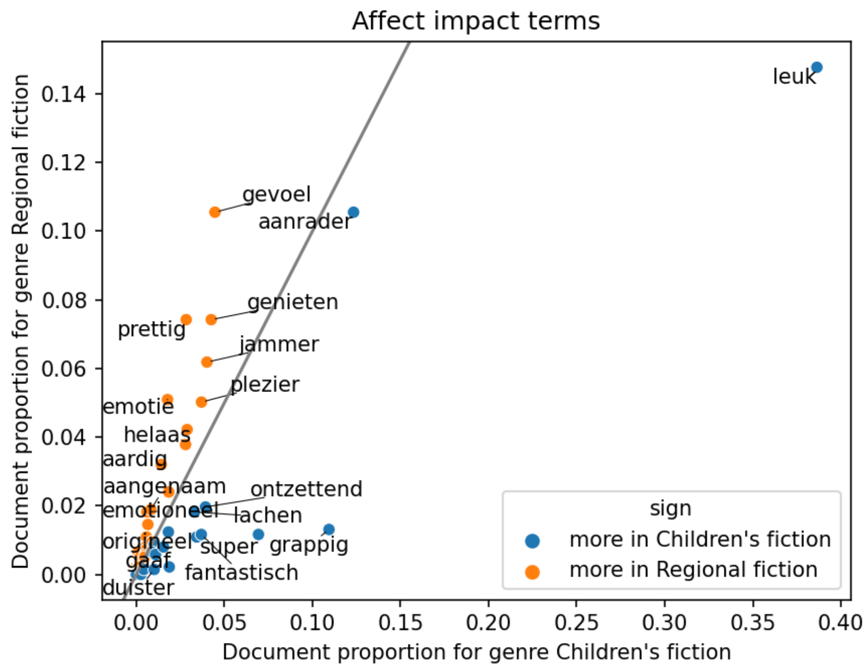


Figure 7: Document proportions of generic *Affect* terms for Children's fiction and Regional fiction.

Vocabulary differences between genres We compute the *Zeta* scores between pairs of genres for all impact terms and sum these scores per impact type to find which pairs of genres have the largest summed difference of *Zeta* scores. For generic *Affect*, Children's fiction is most distinctive as it has high score differences with all other genres. The document proportions for generic *Affect* terms of Children's fiction and *Regional fiction* are shown in Figure 7. The diagonal line shows where terms have equal proportions in both genres. Reviews of children's fiction seem to use a smaller impact vocabulary – almost all document proportions are close to zero – but much higher proportions for the impact term "leuk" (fun or cool). This term is used much less in reviews of other genres

For *Narrative* impact, the biggest summed difference is between Romance and Literary thrillers (see Figure 8). The main differences are found with a handful of terms, "spannend" (thrilling/suspenseful), "spanning" (suspense) and "verrassen" (surprise) are more common in Literary thrillers and "romantisch" (romantic) and "heerlijk" (lovely, wonderful) are more common in Romance novels. These are perhaps somewhat obvious, but show that impact, or at least the language of impact, is related to genre.

For *Aesthetic* impact, the biggest summed difference is between Romance and Historical fiction (see Figure 9). Here, the main differences are again with a few terms. Reviews of Historical fiction more often mention impact terms like "mooi" (beautiful), "beschrijven" (describe), "beschreven" (described) and "prachtig" (beautiful). Reviews of Romance novels more often mention "schrijfstijl" (writing style), "humor" (humor) and "luchtig" (airy). It seems that for Historical fiction, reviewers focus more on descriptions (how evocatively the author describes historical settings, persons or events perhaps), while reviewers of Romance novels focus more on humor and lightness of style. A close reading of some of the contexts in which "schrijfstijl" is mentioned in Romance reviews suggest that reviewers often use it in phrases like "makkelijke schrijfstijl" and "vlotte schrijfstijl" (a writing style that reads easily or quickly respectively).

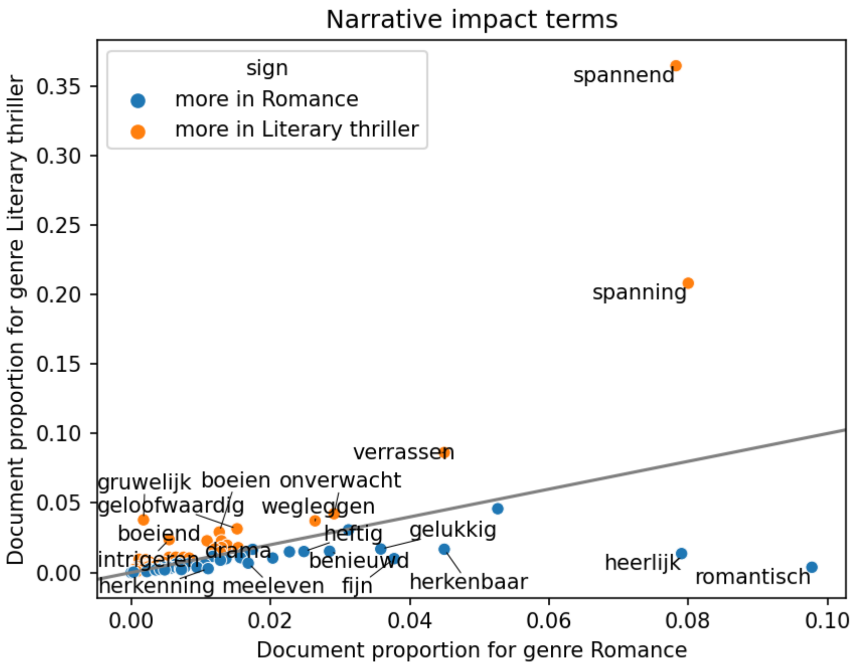


Figure 8: Document proportions of *Narrative* impact terms for Romance and Literary thrillers.

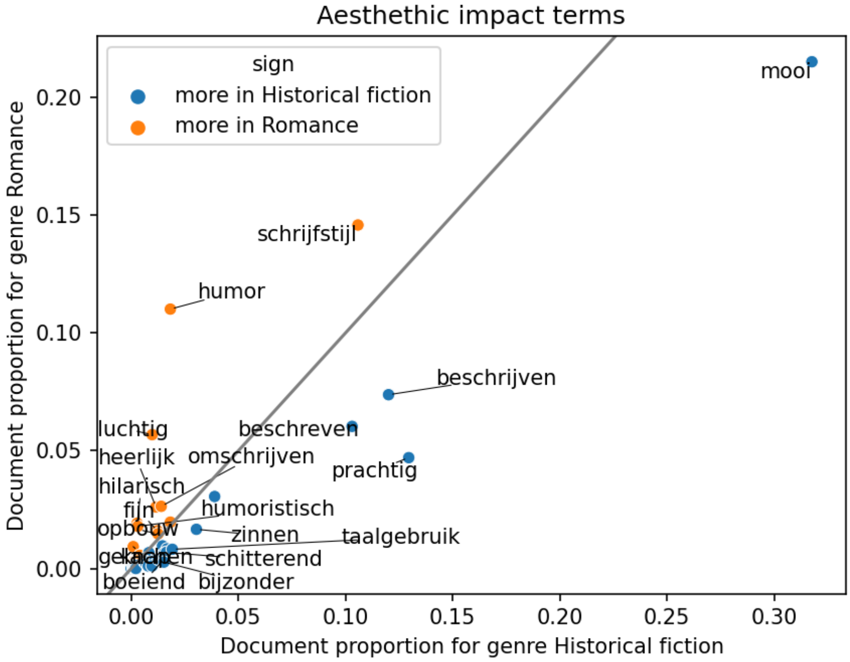


Figure 9: Document proportions of *Aesthetic* impact terms for Historical fiction and Romance.

4.3 Impact and Topic

The third link between the three main concepts that are the focus of this paper is between impact and topic.

To study how the use of impact terms differs between reviews of books with different themes, we first need to group the reviews by theme. Because themes are based on topics and some themes share the same topics, some reviews are assigned to multiple themes. We calculated correlations between themes in terms of the $\%Diff$ per impact term, just as we did for genre (see Figures 10, 11 and 12 in Appendix C). There are many observations that could be made, but again we limit ourselves to the most salient ones related to the three largest themes (in number of books).

Generic Affect

The theme *geography & setting* has a strong correlation for generic Affect with *history* ($\rho = 0.68$) and moderate correlations with *crime* ($\rho = 0.46$) and *war* ($\rho = 0.44$). This is not due to a large overlap in books, as *culture* has the largest overlap with *geography & setting* (sharing 49% and 40% of their books respectively), but a moderately negative correlation ($\rho = -0.41$). With all the other themes, *geography & setting* has no to moderately negative correlations. The connections with *crime*, *history* and *war* make sense, to the extent that for all these themes (we assume), the aspect of place plays an important role. Why this results in similarities of how generic affect is expressed is not immediately clear.

The theme *behaviors / feelings* has moderate correlations for generic Affect with *lifestyle & sport* ($\rho = 0.55$) and *romance & sex* ($\rho = 0.56$). This is partly explained by the latter themes sharing 15% and 22% of their books with *behaviors / feelings*, but it cannot be the only explanation. *Family* shares 65% of its books with *behaviors / feelings* but has no correlation ($\rho = 0.19$).

The theme *culture* has a near perfect correlation with *travel & transport* in terms of generic affect, but no to moderately negative correlations with all other themes. Here the overlap in books is minimal, the two themes sharing respectively 2% and 6% of their books. As mentioned above, With em geography & setting it has a moderately negative correlation ($\rho = -0.41$) despite its substantial overlap.

Narrative Impact

For Narrative impact, the correlations between *geography & setting* are somewhat different. We again find strong and moderate correlations with *history* ($\rho 0.65$) and *war* ($\rho 0.48$) respectively, but also with *religion, spirituality and philosophy* ($\rho 0.46$) and only a weak correlation with *crime* ($\rho 0.30$).

The theme *behaviors / feelings* only has strong correlation with *culture* ($\rho = 0.67$) but no or weakly negative correlations with all others, despite its overlap with *culture* (sharing 13% and 14% of their books respectively) being similar or lower than with *geography & setting* (sharing 13% and 12%) and with *economy & work* (sharing 12% and 36%). Overlap in books is clearly not the main explanation in overlap in the use of impact terms.

The *culture* theme has the strong correlation with *behaviors / feelings* mentioned above, but no or weakly negative correlations with other themes. Again, books with em culture as a theme have a different relationship with how reviewers describe impact than *geography & setting*, despite sharing a substantial number of books.

Aesthetic Impact

For *Aesthetic* impact, *geography & setting* has moderate correlations with *crime* (ρ0.35), *culture* (ρ0.49), *religion, spirituality and philosophy* (ρ0.42) and *war* (ρ0.42). With *crime*, *culture* and *war* this could be due to their substantial overlap in books, but again, overlap cannot but the full explanation, as *geography & setting* also substantially overlaps with *history* while having a moderately negative correlation with it ($\rho - 0.41$).

The *behaviors / feelings* theme has a strong correlation with *romance & sex* ($\rho = 0.71$) and moderate correlations with *family* ($\rho = 0.48$), *lifestyle & sport* ($\rho = 0.45$) and *science* ($\rho = 0.52$), and no or negatively weak correlations with other themes. As mentioned before, 65% of books in the *family* theme also belong to *behaviors / feelings*, but *science* shares no books with *behaviors / feelings*.

Just on these observations alone, it seems that themes have different relationships with how reviewers express the impact of books that cover these themes.

5. Discussion & Conclusion

In this paper we investigated the relationship between three important concepts in literary studies: genre, topic and impact (more commonly known as “reader response”). We discuss our findings for each pair of concepts in turn.

Genre and Topic Our analyzes have corroborated earlier findings on the relationship between genre and topic. By clustering topics identified by topic modelling into broader themes, and measuring the prevalence of these themes in the books of specific genres, we find that topics have a strong relation with genres, and the genres have distinct thematic profiles. These profiles match existing intuitions about the distribution of themes across genres. Potentially these profiles can provide additional insight in genre dynamics (e.g. as to what motivates readers to mix-read genres or not) although much of this aspect remains to be examined.

Genre and Impact The Dutch Reading Impact Model (DRIM, Boot and Koolen 2020) identifies sets of words that are to some extent related to genre, and by studying the overlap in key impact terms between genres, we find clusters of genres that are similar in how their impact is described. Of course, this is not entirely surprising. For instance, *Suspense novels* and *Literary thrillers* are more similar in terms of overall impact. However, it is much less obvious or intuitive that these two genres are more similar in terms of stylistic impact than in terms of narrative impact. Neither is it immediately obvious why literary fiction with respect to all types of impact differs most from other genres.

It remains unclear for now how we should explain the the relationship between impact and genre. Perhaps this relation signals that reviewers develop and copy conventions of writing about books from a certain genre by adopting what others in a genre-related community do. For instance, in a community of reviewers around crime novels and literary thrillers reviewers might converge on a shared vocabulary for talking about the plot and their reading experiences. It could also be that different types of readers are drawn to different types of genres, with each group having their own characteristics that shape how they write their reviews. Another possibility is that reviewers are influenced by the language used by the authors of the novels they read, and how those authors adopt genre conventions. Finally, depending on how the model was developed, this may also be an artifact of how the rules were constructed. For instance, if reviews per genre were scanned to identify common expressions of impact. Further analysis is required to establish which, if any, of these factors contributes to the relationship between fiction genres and reading impact as expressed in reviews.

Topic and Impact For the first two pairs of concepts, there were some expectations, e.g. that there is a relation between the Romance genre and topics related to the theme of *Romance and sex*, or that typical narrative impact terms in reviews of Young adult novels overlap with those in reviews of Fantasy novels. For the link between topic and impact, we struggled to come up in advance with expectations on how the topics in novels are related to impact. Novels discussing topics such as war and its consequences or living with physical or mental illness might lead to more reviews mentioning narrative impact. But honest reflection forces us to admit that the results of topic modelling are still far removed from explaining how authors deal with topics and how reviewers discuss them. This remove stubbornly persists throughout continued engagements with our data in several papers. This should give us pause to reflect on our operationalizations that are by and large still based on bags-of-words approach. Vector modelings are becoming increasingly more sophisticated. Nevertheless we have not inched significantly closer to answering the question what features of novel texts relate to what types of reader impact adequately and satisfyingly from a literary studies perspective.

Our reflections tie in with observations and suggestions made in some recent methodological publications on computational humanities. Bode (2023) argues that humanities researchers applying conventional methods and those that embrace computational or data-science methods should take a greater and more sincere interest in each others' work. Rather than addressing research questions by stretching either method beyond limits, researchers ought to investigate how the different methods can reinforce and amplify each other. Pichler and Reiter (2022) argue that operationalizations in computational linguistics and computational literary studies are currently often poor because we typically fail to express the precise operations that identify the theoretical concept we are trying to observe. Indeed our operationalizations seem underwhelming in the light of literary mechanisms. The reason to label a topic as being *about* war is that it contains words directly and strongly associated with war, and emphasizing the physical aspects of it, such as *war*, *soldier*, *bombing*, *battlefield*, *wounded*, etc. But novels that readers would describe as being *about* war might instead focus on more indirect aspects or on aspects that war shares with many other situations, such as dire living conditions or being cut-off from the rest of the world, feeling unsafe and scared, or the sense of helplessness or hopelessness. And it is not just that war-related words to describe these aspects might lead an annotator to label a topic as being about something other than war. It is also that an author, going by the good practice of "show don't tell" can conjure up images that fit these words in almost infinitely many ways that are almost impossible to capture by looking at bags of words. Which means we need infinitely better operationalizations.

6. Data Availability

Data used for the research can be found at: https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact.

7. Software Availability

All code created and used in this research has been published at: https://github.com/impact-and-fiction/jcls-2024-topic-genre-impact.

8. Acknowledgements

This project has been supported through generous material and in-kind technical and data-science analytical support from the eScience Center in Amsterdam. We thank the National Library of the Netherlands for providing access to the novels used in this research and for their invaluable technical support.

9. Author Contributions

Marijn Koolen: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft

Joris J. Van Zundert: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Visualization, Writing – review & editing

Eva Viviani: Formal analysis, Software, Validation, Visualization

Carsten Schnober: Resources, Software

Willem Van Hage: Methodology, Resources, Software

Katja Tereshko: Writing – original draft, Writing – review & editing

References

Abrams, M.H. (1971). *The Mirror and the Lamp: Romantic Theory and the Critical Tradition*. Oxford etc.: Oxford University Press.

Angelov, Dimo (2020). “Top2vec: Distributed representations of topics”. In: *arXiv preprint arXiv:2008.09470*.

Ashok, Vikas Ganjigunte, Song Feng, and Yejin Choi (2013). “Success with Style: Using Writing Style to Predict the Success of Novels”. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington: Association for Computational Linguistics, 1753–1764. https://api.semanticscholar.org/CorpusID:7100691 (visited on 07/28/2023).

Berger, Jonah and Grant Packard (2018). “Are Atypical Things More Popular?” In: *Psychological Science* 29.7, 1178–1184. 10.1177/0956797618759465.

Berry, David M. (2014). *Critical Theory and the Digital*. Critical Theory and Contemporary Society. New York, London, New Delhi etc.: Bloomsbury Academic.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). “Latent dirichlet allocation”. In: *Journal of machine Learning research* 3.Jan, 993–1022.

Bode, Katherine (2023). "What's the Matter with Computational Literary Studies?" In: *Critical Inquiry* 49.4, 507–529.

Boot, Peter (2017). "A Database of Online Book Response and the Nature of the Literary Thriller". In: *Digital Humanities*, 4.

Boot, Peter and Marijn Koolen (2020). "Captivating, splendid or instructive?: Assessing the impact of reading in online book reviews". In: *Scientific Study of Literature* 10.1, 35–63. https://www.jbe-platform.com/content/journals/10.1075/ssol.20003.boo (visited on 01/22/2024).

Burrows, John (2006). "All the way through: testing for authorship in different frequency strata". In: *Literary and Linguistic Computing* 22.1, 27–47.

Cranenburgh, Andreas van, K.H. van Dalen-Oskam, and Joris J. van Zundert (2019). "Vector space explorations of literary language". In: *Language Resources and Evaluation* 53.4, 625–650. 10.1007/s10579-018-09442-4.

Culpeper, Jonathan and Jane Demmen (2015). "Keywords". In: *The Cambridge handbook of English corpus linguistics*, 90–105.

Darbyshire, Madison (2023). "Hot stuff: why readers fell in love with romance novels". In: *Financial Times*. https://www.ft.com/content/0001f781-4927-4780-b46c-3a9f15dffe78 (visited on 01/22/2024).

Du, Keli, Julia Dudar, Cora Rok, and Christof Schöch (2021). "Zeta & eta: An exploration and evaluation of two dispersion-based measures of distinctiveness". In: *Proceedings http://ceur-ws.org ISSN* 1613, 0073.

Du, Keli, Julia Dudar, and Christof Schöch (2022). "Evaluation of measures of distinctiveness. Classification of literary texts on the basis of distinctive words". In: *Journal of Computational Literary Studies* 1.1.

Dunning, Ted (1994). "Accurate methods for the statistics of surprise and coincidence". In: *Computational linguistics* 19.1, 61–74.

Fialho, Olivia (2019). "What is literature for? The role of transformative reading". In: *Cogent Arts & Humanities* 6.1. Ed. by Anezka Kuzmicova, 1692532. 10.1080/23311983.2019.1692532.

Gabrielatos, Costas (2018). "Keyness analysis". In: *Corpus approaches to discourse: A critical review*, 225–258.

Gitelman, Lisa, ed. (2013). *"Raw Data" Is an Oxymoron*. Cambridge: The MIT Press.

Gries, Stefan Th (2008). "Dispersions and adjusted frequencies in corpora". In: *International journal of corpus linguistics* 13.4, 403–437.

Hickman, Miranda B. (2012). "Introduction: Rereading the New Criticism". In: *Rereading the New Criticism*. Ed. by Miranda B. Hickman and John D. McIntyre. Columbus: The Ohio State University Press, 1–21. http://hdl.handle.net/1811/51698 (visited on 01/16/2024).

Koolen, Marijn, Peter Boot, and Joris J van Zundert (2020). "Online Book Reviews and the Computational Modelling of Reading Impact". In: *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)*. Vol. 2723, 0073. http://ceur-ws.org/Vol-2723/long13.pdf.

Koolen, Marijn, Olivia Fialho, Julia Neugarten, Joris J. van Zundert, Willem van Hage, Ole Mussmann, and P. Boot (2023). "How Can Online Book Reviews Validate Empirical In-depth Fiction Reading Typologies?" In: *IGEL 2023 : Rhythm, Speed, Path: Spatiotemporal Experiences in Narrative, Poetry, and Drama*. Monopoli: NARRNET, IGEL, elit, 1. https://igel society.org/events/igel2023/#submission-requirements (visited on 01/16/2024).

Laurino Dos Santos, Henrique and Jonah Berger (2022). “The speed of stories: Semantic progression and narrative success.” In: *Journal of experimental psychology. General* 151.8, 1833–1842. 10.1037/xge0001171.

Lijffijt, Jefrey, Terttu Nevalainen, Tanja Säily, Panagiotis Papapetrou, Kai Puolamäki, and Heikki Mannila (2016). “Significance testing of word frequencies in corpora”. In: *Digital Scholarship in the Humanities* 31.2, 374–397.

Loi, C., F. Hakemulder, M. Kuijpers, and G. Lauer (2023). “On how Fiction Impacts the Self-Concept: Transformative Reading Experiences and Storyworld Possible Selves”. In: *Scientific Study of Literature* 12.1, 44–67. 10.61645/ssol.181.

Manovich, Lev (2013). *Software Takes Command*. Vol. 5. International Texts in Critical Media Aesthestics. New York, London, New Delhi etc.: Bloomsbury Academic.

Miall, David S. and Don Kuiken (1994). “Beyond text theory: Understanding literary response”. In: *Discourse processes* 17.3, 337–352. 10.1080/01638539409544873.

Moreira, Pascale, Yuri Bizzoni, Kristoffer Nielbo, Ida Marie Lassen, and Mads Thomsen (2023). “Modeling Readers’ Appreciation of Literary Narratives Through Sentiment Arcs and Semantic Profiles”. In: *Proceedings of the The 5th Workshop on Narrative Understanding*. Toronto, Canada: Association for Computational Linguistics, 25–35. https://aclanthology.org/2023.wnu-1.5 (visited on 01/22/2024).

Nguyen, Minh Van, Viet Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen (2021). “Trankit: A Light-Weight Transformer-based Toolkit for Multilingual Natural Language Processing”. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*.

Paquot, Magali and Yves Bestgen (2009). “Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction”. In: *Corpora: Pragmatics and discourse*. Brill, 247–269.

Pichler, Axel and Nils Reiter (2022). “From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities”. In: *Journal of Cultural Analytics* 7.4.

Prescott, Andrew (2023). “Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities?” In: *DHQ: Digital Humanities Quarterly* 17.2, 689. https://www.digitalhumanities.org/dhq/vol/17/2/000689/000689.html (visited on 01/16/2024).

Rawson, Katie and Trevor Muñoz (July 2016). *Against Cleaning*. Project blog. http://curatingmenus.org/articles/against-cleaning/ (visited on 09/30/2016).

Regis, Pamela (2003). *A Natural History of the Romance Novel*. Philadelphia: University of Pennsylvania Press.

Schöch, Christof (2017). “Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama.” In: *DHQ: Digital Humanities Quarterly* 11.2.

Sobchuk, Oleg and Artjoms Šeļa (2023). “Computational thematics: Comparing algorithms for clustering the genres of literary fiction”. In: *arXiv preprint arXiv:2305.11251*.

Thompson, Laure and David Mimno (2018). “Authorless Topic Models: Biasing Models Away from Known Structure”. In: *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 3903–3914. https://aclanthology.org/C18-1329 (visited on 01/22/2024).

Toubia, Olivier, Jonah Berger, and Jehoshua Eliashberg (2021). “How quantifying the shape of stories predicts their success”. In: *Proceedings of the National Academy of Sciences* 118.26, 1–5. 10.1073/pnas.2011695118.

Uglanova, Inna and Evelyn Gius (2020). “The Order of Things. A Study on Topic Modelling of Literary Texts.” In: *CHR* 18-20, 2020.

Van Zundert, Joris, Marijn Koolen, Julia Neugarten, Peter Boot, Willem Van Hage, and Ole Mussmann (2022). “What Do We Talk About When We Talk About Topic?” In: *Proceedings of the Conference on Computational Humanities Research 2022*. Antwerpen: CEUR Workshop Proceedings, 398–410. https://ceur-ws.org/Vol-3290/ (visited on 11/22/2023).

Van Zundert, Joris, Marijn Koolen, and Karina Van Dalen-Oskam (2018). “Predicting Prose that Sells: Issues of Open Data in a Case of Applied Machine Learning”. In: *JADH 2018 ”Leveraging Open Data”: Proceedings of the 8th Conference of Japanese Association for Digital Humanities*. Tokyo: Center for Open Data in the Humanities, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 175–177. https://conf2018.jadh.org/files/Proceedings_JADH2018_rev0911.pdf (visited on 11/07/2018).

Warnock, John (1978). “A THEORY OF DISCOURSE, by James L. Kinneavy. (Review)”. In: *Style* 12.1. Ed. by James L. Kinneavy. Publisher: Penn State University Press, 52–54. http://www.jstor.org.proxy.uba.uva.nl/stable/45109026 (visited on 01/16/2024).

Wimsatt, W.K. (1954). “The Intentional Fallacy”. In: *The Verbal Icon: Studies in the Meaning of Poetry*. and two essays written in collaboration with Monroe C. Beardsley. Lexington: The University Press of Kentucky, 3–20.

NUR code	NUR label	Genre label
280	Children's Fiction general	Children's fiction
281	Children's fiction 4 - 6 years	Children's fiction
282	Children's fiction 7 - 9 years	Children's fiction
283	Children's fiction 10 - 12 years	Children's fiction
284	Children's fiction 13 - 15 years	Young adult
285	Children's fiction 15+	Young adult
300	Literary fiction general	Literary fiction
301	Literary fiction Dutch	Literary fiction
302	Literary fiction translated	Literary fiction
305	Literary thriller	Literary thriller
312	Pockets popular fiction	Literary fiction
313	Pockets suspense	Suspense
330	Suspense general	Suspense
331	Detective	Suspense
332	Thriller	Suspense
334	Fantasy	Fantasy fiction
339	True crime	Suspense
342	Historical novel (popular)	Historical fiction
343	Romanticism	Romanticism
344	Regional- and family novel	Regional fiction

Table 2: The selected NUR codes of novels in our dataset of 18,885 novels, and their mapping to genres.

A. Mapping NUR Codes to Genre Labels

The complete mapping from NUR codes to genre labels is shown in Table 2.

B. Overlap between Themes in Terms of Shared Books

The topic modelling process assigns each book to a single topic, but because individual topics can linked to multiple themes, their books are also linked to multiple themes. As a consequence, themes share books and reviews and some pairs of themes may have larger overlap than others. This overlap between themes is shown for pairs of themes where for one theme, at least 25% of books are shared by the other theme.

C. Correlations between Themes in Terms of Impact

The correlations between themes in terms of the percent difference (%Diff) per impact term for generic *Affect*, *Narrative* and *Aesthetics* is shown respectively in Figures 10, 11 and 12.

Theme 1	Share 1	Theme 2	Share 2	Book overlap	Books theme 1	Books theme 2
crime	0.33	geo. & setting	0.14	619	1899	4317
culture	0.49	geo. & setting	0.40	1713	3524	4317
econ. & work	0.36	behav./feelings	0.12	446	1232	3860
econ. & work	0.30	society	0.44	371	1232	851
econ. & work	0.25	politics	0.49	310	1232	634
family	0.65	behav./feelings	0.08	324	498	3860
family	0.30	culture	0.04	151	498	3524
geo. & setting	0.40	culture	0.49	1713	4317	3524
history	0.51	geo. & setting	0.24	1038	2020	4317
history	0.31	war	0.65	622	2020	952
lifest. & sport	0.31	medi./health	0.20	216	702	1058
politics	0.49	econ. & work	0.25	310	634	1232
politics	0.49	society	0.36	310	634	851
society	0.44	econ. & work	0.30	371	851	1232
society	0.36	politics	0.49	310	851	634
war	0.65	history	0.31	622	952	2020

Table 3: Overlap in books between themes, for themes where one theme shares at least 25% of books with the other theme.

Figure 10: Percent different correlations between themes based on general *Affect* terms.

Figure 11: Percent different correlations between themes based on *Narrative* impact terms.

Figure 12: Percent different correlations between themes based on general *Aesthetic* impact terms.